



Localization of broadband sounds carrying interaural time differences: Effects of frequency, reference location, and interaural coherence

Buchholz, Jorg M.; Le Goff, Nicolas; Dau, Torsten

Published in:

Journal of the Acoustical Society of America

Link to article, DOI:

[10.1121/1.5058776](https://doi.org/10.1121/1.5058776)

Publication date:

2018

Document Version

Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):

Buchholz, J. M., Le Goff, N., & Dau, T. (2018). Localization of broadband sounds carrying interaural time differences: Effects of frequency, reference location, and interaural coherence. *Journal of the Acoustical Society of America*, 144(4), 2225–2237. DOI: 10.1121/1.5058776

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Localization of broadband sounds carrying interaural time differences: Effects of frequency, reference location, and interaural coherence

Jörg M. Buchholz^{a)}

Department of Linguistics—Audiology Group, Macquarie University, Australian Hearing Hub,
16 University Avenue, Sydney NSW 2109, Australia

Nicolas Le Goff^{b)} and Torsten Dau

Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark,
DTU Bygning 352, Ørstedes Plads, 2800 Kgs. Lyngby, Denmark

(Received 17 December 2017; revised 9 September 2018; accepted 19 September 2018; published online 16 October 2018)

The auditory processes involved in the localization of sounds in rooms are still poorly understood. The present study investigated the auditory system's across-frequency processing of interaural time differences (ITDs) and the impact of the interaural coherence (IC) of the stimuli in ITD discrimination and localization. First, ITD discrimination thresholds were measured as a function of signal frequency, reference ITD, and IC using critical-band wide noises. The resulting data were fitted with a set of analytical functions and ITD weights were derived using concepts from signal detection theory. Inspired by the weighted-image model [Stern, Zeiberg, and Trahiotis. (1988). *J. Acoust. Soc. Am.* **84**, 156–165], the derived ITD weights were then integrated in a simplified localization model using an optimal combination of ITD information across frequency. To verify this model, a series of localization experiments were conducted using broadband noise in which ITD and IC were varied across frequency. The model predictions were in good agreement with the experimental data, supporting the assumption that the auditory system performs a weighted integration of ITD information across frequency to localize a sound source. The results could be valuable for the design of new paradigms to measure localization in more complex acoustic conditions and may provide constraints for future localization models.

© 2018 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.1121/1.5058776>

[JB]

Pages: 2225–2237

I. INTRODUCTION

In daily reverberant environments, people are not only exposed to sound that travels directly from the source to their ears, but also to the sound reflected from surrounding surfaces. Sound source localization can be challenged because the reflections carry spatial cues, such as interaural time differences (ITDs) and interaural level differences (ILDs), which do not directly correspond to the true source location. Reverberation in rooms does not affect all ITDs and ILDs carried by the sound to the same degree. ITDs and ILDs at the signal onsets are predominantly driven by the direct sound and are less affected by reverberation than ITDs and ILDs carried by the steady-state portions of the signal. Within the steady-state portions, the direct sound and reflections overlap in time which leads to a decrease of the interaural correlation of the ear signals, relative to an anechoic condition where only the direct sound is present. The interaction of the direct sound and its reflections results in

variations of the ITDs, ILDs and the interaural coherence (IC) as a function of time and frequency (Blauert, 1986; Kuttruff, 2000; Kopčo and Shinn-Cunningham, 2002; Hartmann *et al.*, 2005; Westermann *et al.*, 2013), with the IC reflecting the maximum of the normalized cross-correlation function of the left- and right-ear signals (e.g., Faller and Merimaa, 2004).

The auditory system is known to utilize the robustness of the ITDs and ILDs carried by the onsets to successfully localize sounds in reverberant environments. This ability has been associated with the precedence effect (Wallach *et al.*, 1949), an auditory mechanism that emphasizes the spatial cues of the first-arriving wavefront (i.e., the direct sound) and suppresses the spatial cues carried by reflections (see Litovsky *et al.*, 1999, for a review). Rakerd and Hartmann (2005) investigated the importance of the signal's onset for localization as a function of the amount of reverberation. They demonstrated that the preservation of the signal's onset improved the listeners' localization performance, particularly in strongly reverberant conditions, whereas in moderately reverberant conditions, the ITDs and ILDs carried in the steady-state portions already lead to accurate localization results. Stecker and Moore (2018) measured the temporal variation of auditory sensitivity to sound-localization cues in click trains and observed an

^{a)}Also at: The HEARing Cooperative Research Centre, 550 Swanston Street, Carlton VIC 3010, Australia. Electronic mail: jorg.buchholz@mq.edu.au

^{b)}Also at: Oticon A/S, Kongebakken 9, 2765 Smørum, Denmark.

increased perceptual weight of the initial click and a reduced weight of the later clicks in a (simulated) reverberant condition when compared to an anechoic condition. While the importance of the signal's onset for sound localization in reverberant environments has been considered in various investigations (see also Litovsky *et al.*, 1999; Blauert, 1997), the present study examined how the perception of the ITDs in the steady-state portions of a signal is affected by reverberation.

It has been shown that the listeners' sensitivity to ITDs is reduced for signals with a reduced IC at the listeners' ears (e.g., Jeffress *et al.*, 1962; Rakerd and Hartmann, 2010), suggesting that sound localization performance also decreases with decreasing IC. At the same time, a reduction in IC may be perceived as a broadening of the apparent source width or as an increased sense of being immersed or enveloped in the sound (ISO 3382-1, 2009). Faller and Merimaa (2004) presented a model framework for predicting the localization of multiple sound sources in anechoic as well as reverberant environments. This model includes a "cue-selection" mechanism whereby instantaneous ITDs and ILDs are estimated as reliable (for localization) when the instantaneous IC is above a predefined threshold. This mechanism is also included in the model of Le Goff *et al.* (2013a), which is based on the equalization cancellation approach (Durlach, 1963), as well as the binaural multi-source localization model proposed by Dietz *et al.* (2011). As discussed in Faller and Merimaa (2004) and Le Goff *et al.* (2013a), a shortcoming of the cue-selection mechanism is that the IC threshold is chosen arbitrarily and that the best predictions are obtained for IC thresholds that depend on frequency and the amount of room reverberation. Furthermore, neither model specifies effects of integration of ITD and ILD information across frequency, although most natural sounds, such as speech, are broadband. Kayser *et al.* (2015) addressed some of these limitations by applying a probabilistic model as a back-end to the binaural model proposed by Dietz *et al.* (2011) using an IC-based weighting of the interaural cues. Even though this model provided robust localization performance in different complex acoustic environments, it provided a rather technical solution with only limited psychoacoustic relevance.

To account for effects of spectral integration in localization, Stern *et al.* (1988) proposed the "weighted-image model" as a conceptual extension to existing cross-correlation-based localization models. In their approach, it is assumed that the input signals are first decomposed into frequency channels (cochlear filters) and that the internal representation of the ITD in each frequency channel is weighted before information is (linearly) combined across frequency. The weighting is achieved with three components. The first component is termed "centrality" and emphasizes the internal representations of the ITDs corresponding to sound source locations close to the median plane. The second component provides a bandpass-filter shaped weighting with emphasis around 600 Hz, based on experimental data obtained in Raatgever (1980). The third component has been termed "straightness" and provides a weight de-emphasis when the ITD values in adjacent frequency channels are not equal, i.e., not "straight." The weighted-image model was evaluated in Stern *et al.*

(1988) and Trahiotis and Stern (1989) by comparing model predictions to a large set of localization data obtained with different types of low-frequency bandpass filtered stimuli. The origin of the applied weighting functions and their parameters are described in Stern and Shear (1998). Shackleton *et al.* (1992) presented a simplified and physiologically more plausible version of this model.

A major limitation of the study of Stern *et al.* (1988) is that the effects of room reverberation are not considered. Specifically, the weight of the ITD information may decrease with decreasing IC at the listener's ears in a frequency-dependent way (Faller and Merimaa, 2004; LeGoff *et al.*, 2013a). Since the weighted-image model only considers the location of the maxima of the (long-term) cross-correlation function, it is not sensitive to the height of the cross-correlation function which is linked to the IC. Furthermore, the model does not consider other processes that have been essential in most existing binaural models (e.g., Colburn, 1977; Cai *et al.*, 1998; Lindemann, 1986; Gaik, 1993; Breebaart *et al.*, 2001; Dietz *et al.*, 2011) and affect the weighting of ITD information across frequency in the back end of the respective models. For example, the contra-lateral inhibition mechanism proposed by Lindemann (1986) and Gaik (1993) affect the amplitude of the estimated binaural (cross-correlation) output depending on the given reference-ITD, ILD, and IC. However, the processes in these models are nonlinear and relatively complex and it is unclear to what extent the different approaches correctly reflect how the auditory system weights ITDs across frequency. In fact, no psychoacoustical data are available that allow the verification (and optimization) of the ITD weighting applied by such models, particularly in conditions with different IC. Such data seem crucial for a better understanding of the processes underlying auditory localization of broadband signals in reverberant conditions.

The goals of the present study were twofold. First, the effect of a reduced IC on the listeners' sensitivity to ITDs was investigated. ITD discrimination thresholds were measured for critical-band wide noises as a function of the center frequency of the noise. This was done for different values of both the presented IC and the reference ITD. The ITD discrimination data were then described using a set of analytical functions. Based on these functions, using concepts from signal detection theory, the variance of the internal "auditory noise" that limits the listeners' discrimination performance was estimated. The variance of this noise term was then used to derive the weight of the ITD information in a given frequency channel as a function of the signal's IC and ITD.

Second, the effect of the IC-dependent ITD sensitivity on sound source localization was studied. The localization performance was measured for bandpass noise in which the ITD and the IC were modified independently in individual frequency channels. A functional sound localization model was developed, inspired by the framework of Stern *et al.* (1988), incorporating the ITD weights derived from the ITD discrimination data of the first experiment. The model was then validated using experimental data on localization of broadband signals carrying frequency-specific ITD and IC.

II. ITD DISCRIMINATION

A. Rationale

To better understand the effect of changes in the IC on localization as, for example, introduced by changes in the amount of room reverberation, three ITD discrimination experiments were conducted. ITD discrimination thresholds were measured as a function of center frequency, reference ITD, and IC for critical-band wide noises placed at different positions along the lateral axis. The resulting threshold functions were approximated by a set of analytical functions which provides input to the localization modeling described in Sec. III C.

B. Method

1. Listeners

Seven young listeners (25–35 yrs) participated in this part of the study, but only four listeners participated in each individual experiment (see Tables I and II) due to the very time consuming testing. The listeners had no evidence or history of hearing loss and were trained for 1–3 h, depending on their experience with the task. One of the listeners was the second author.

2. Apparatus and stimuli

The listeners were seated in a sound-attenuated listening booth in front of a computer screen and a keyboard. All thresholds were measured using a MATLAB program running on a computer equipped with a RME DIGI96 sound card (Audio AG, Am Pfanderling 60, 85778 Haimhausen, Germany). Sennheiser HD 580 headphones were used to present the stimuli, calibrated with a 1-kHz pure tone on a Bruel and Kjaer 4152 artificial ear (Skodsborgvej 307, 2850 Naerum, Denmark).

All noise signals were digitally generated with a sampling rate of 44.1 kHz. Prior to each measurement, a 5-s buffer of bandlimited noise was generated. The buffer was created from a white Gaussian noise in the time domain that was filtered to the desired bandwidth in the frequency domain. The stimuli were presented at a sound pressure level (SPL) of 70 dB. For each interval, a new noise token was generated by randomly selecting a 300-ms portion of the noise buffer that was gated with 5-ms long cosine-shaped onset and offset ramps. The noise token was bandpass-filtered in the frequency domain by setting the amplitude of all frequency bins outside the passband to zero. Ongoing ITDs were created by an all-pass filter that had a constant group delay corresponding to the desired ITD. The filter was realized by applying a phase shift specific to the ITD and to each frequency bin in the spectral domain. The resulting signals at the left and right ear had the same envelope but the fine structure was shifted according to the applied ITD.

ITD discrimination thresholds were measured using bandpass-filtered noise with a bandwidth of one equivalent rectangular bandwidth (ERB; Glasberg and Moore, 1990) that depended on the center frequency of the noise. In the first experiment, this was done for four different values of

the IC: 1, 0.97, 0.92, or 0.85. For the fully correlated signals (IC = 1), thresholds were measured at the center frequencies 148, 231, 330, 451, 498, 776, 992, 1254, and 1572 Hz. The partially coherent stimuli were generated using the symmetric-two-generator method described in Hartmann and Cho (2011). Here, the ITD thresholds were measured at a subset of the center frequencies: 231, 451, 776, and 1254 Hz. The reference ITD was always 0 μ s in this experiment.

In the second and third experiments, the reference ITD was either 200, 400, or 600 μ s. Thresholds were measured at the center frequencies 148, 231, 330, 451, 498, 776, 992, 1254, and 1572 Hz for each of the three reference ITDs. The stimuli were either fully correlated (IC = 1, experiment 2) or had an IC = 0.92 (experiment 3). In these experiments, the reference ITD was applied on the stimuli in all three intervals (as a lateralization to the right side) and was kept constant during each threshold measurement. The target ITD was subtracted from the reference ITD in one of the three randomly selected intervals.

3. Procedure

ITD thresholds were obtained using an adaptive, three-interval, three-alternative forced-choice (3-AFC) procedure in conjunction with a 1-up, 2-down tracking rule to estimate the 70.7% correct point of the psychometric function (Levitt, 1971). Listeners responded via the computer keyboard after each trial whereby no feedback was provided. The initial value of the target ITD, which was subtracted from the reference ITD, was chosen such that all subjects could easily discriminate the lateralization of the target stimulus from the reference stimulus, and varied between 200 and 400 μ s dependent on the considered frequency as well as the applied IC. The initial step size of the adaptive track corresponded to a factor of 1.6 (2 dB) and was reduced to a factor of 1.1 (0.5 dB) after two reversals. The pause between successive intervals was 500 ms. Each run was terminated after ten reversals, and thresholds were defined as the geometric mean over the last eight reversals. Three repetitions of the threshold measurements were made for each subject and for each experiment.

4. Functional description of measured threshold functions

The obtained ITD threshold functions were approximated by analytical functions. According to the concept of signal detection theory (Green and Swets, 1966), the variance of the noise term, σ^2 , that limits the discrimination performance, is related to the measured ITD thresholds Δ_τ by

$$\sigma^2(f_0, \tau, \rho) = \left(\frac{\Delta_\tau(f_0, \tau, \rho)}{d'} \right)^2, \quad (1)$$

with ρ representing the applied IC, f_0 the center frequency of the 1-ERB-wide noise, d' the sensitivity index defined by the applied experimental method, and τ the considered ITD [see Appendix A, Eq. (A2)]. For the 3-AFC task applied in the discrimination experiments to measure the 70.7% point on

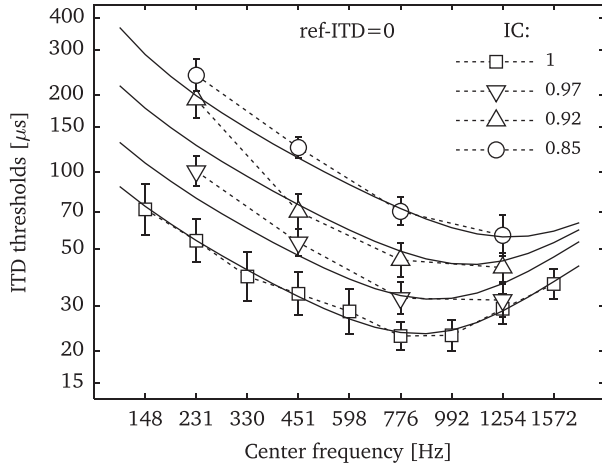


FIG. 1. ITD discrimination thresholds for 1-ERB-wide Gaussian noise measured as a function of its center frequency. The reference ITD was equal to 0 μ s and the parameter was the IC which was either 1, 0.97, 0.92, or 0.85. The data represent the mean thresholds of the four listeners. Error bars represent the 95% confidence interval of the mean. The continuous lines represent the fitted function to the data.

the psychometric function, d' corresponds to a value of 1.28 (Hacker and Ratcliff, 1979).

As in Bernstein and Trahiotis (2008), it was further assumed that the noise term can be divided in two components,

$$\sigma^2(f_0, \tau, \rho) = \sigma_{\text{int}}^2(f_0, \tau) + \sigma_{\text{ext}}^2(f_0, \tau, \rho), \quad (2)$$

whereby the first noise variance component, σ_{int}^2 , reflects an “internal” source of variability that characterizes the limit of the hearing system to code ITDs and is independent of the properties of the physical stimulus. The second noise variance component, σ_{ext}^2 , represents an external source of variability which characterizes the variability of the interaural properties of the physical stimulus.

The fitting of the ITD threshold functions was achieved by first calculating the variance σ^2 for all the measured ITD

thresholds using Eq. (1) and then comparing the results to the corresponding predicted variances using Eq. (2). The variances σ_{int}^2 and σ_{ext}^2 in Eq. (2) were represented in the analytical functions described in Appendix A and fitted to the data by minimizing the mean squared error between the measured and predicted variances. The fitting procedure included some constraints regarding the values of σ_{int}^2 and σ_{ext}^2 depending on the physical properties of the signals (i.e., center frequency, IC, and ITD) and made assumptions in relation to properties of auditory signal processing (phase locking, hair-cell transduction, cochlear filtering), as specified in Appendix A. This approach was found to describe the behavior of the experimental data more accurately than more common approaches (e.g., using multi-dimensional splines or polynomials). The obtained fitted functions are indicated by the solid lines in Figs. 1 and 2.

C. Results and discussion

The results of the first experiment are shown in Fig. 1. The average thresholds across listeners obtained with fully correlated signals (IC = 1), indicated by the squares, decrease with increasing center frequencies up to 776 Hz, and increase above 992 Hz with further increasing center frequency. The threshold values and their frequency dependence are consistent with ITD thresholds obtained with tones (e.g., Klumpp and Eady, 1956; Zwislocki and Feldman, 1956; Brughera *et al.*, 2013). The range between about 750 and 1000 Hz, where ITD thresholds are at a minimum, has sometimes been referred to as the “dominance region,” although the reported frequency range is typically around 600 Hz (e.g., Raatgever, 1980). At low frequencies, the decrease with increasing frequency is roughly linear, consistent with Moore (2012, p. 251), and resembles a sensitivity threshold that corresponds to a constant interaural phase change. The observation that ITD thresholds could not be measured reliably for IC < 1 at the highest considered center frequency of 1572 Hz

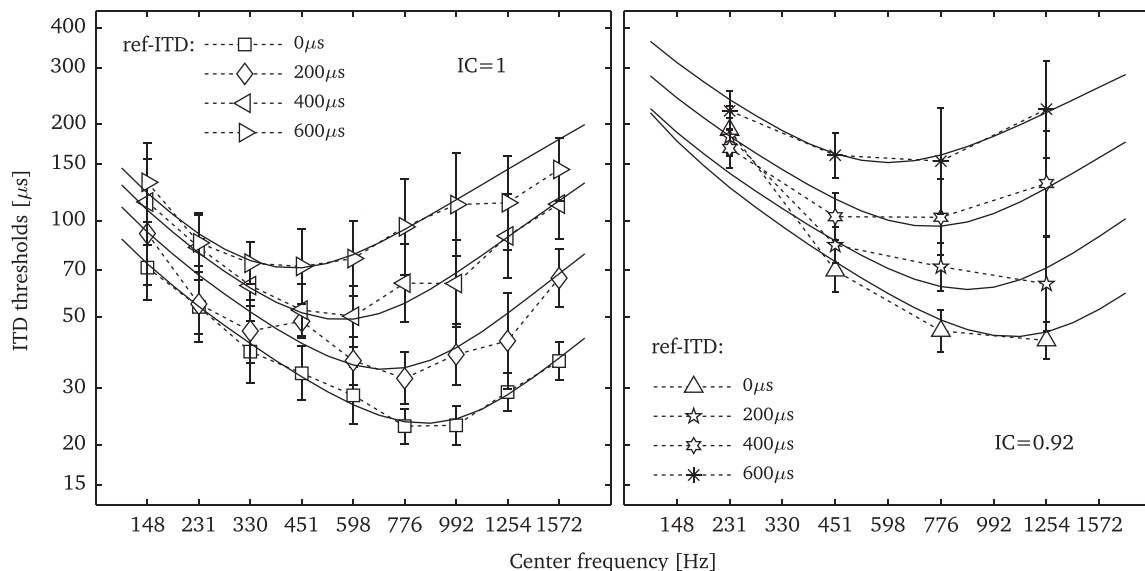


FIG. 2. ITD discrimination thresholds for 1-ERB-wide Gaussian noise as a function of its center frequency, for IC = 1 (left panel) and IC = 0.92 (right panel). In both panels, the parameter was the reference ITD, which was either 0, 200, 400, or 600 μ s. The data, connected by dotted lines, represent the mean thresholds for four listeners. Error bars represent the 95% confidence interval of the mean. The continuous lines represent the fitted function to the data.

is consistent with [Brughera et al. \(2013\)](#) who reported a rapid roll-off of the auditory sensitivity to ITDs for tones above 1000 Hz with unmeasurable thresholds just above 1400 Hz. In this regard, the measured ITD threshold at $IC = 1$ of $37.5 \mu s$ may be surprising. However, the narrowband noise stimulus, with its lower -3 dB cutoff frequency of 1480 Hz and limited frequency roll-off (Sec. II B 2), may have still provided sufficient stimulus energy below 1400 Hz for the auditory system to evaluate ITDs.

The thresholds obtained with the partially correlated noise (downward triangles, $IC = 0.97$; upward triangles, $IC = 0.92$; circles, $IC = 0.85$) are above those obtained with the fully correlated noise. This is consistent with the data from previous studies obtained with broadband signals (e.g., [Jeffress et al., 1962](#); [Rakerd and Hartmann, 2010](#)). The size of the increase of the ITD thresholds with decreasing IC depends on the center frequency of the noise. For example, for the low-frequency noise centered at 231 Hz, the threshold obtained for $IC = 0.85$ is 4.3 times larger than that obtained with the fully correlated noise. At the center frequency of 1254 Hz, the corresponding ratio is only 2.0.

The results obtained in the second experiment are shown in the left panel of Fig. 2. The thresholds for the reference ITD of $0 \mu s$ (squares) were replotted from Fig. 1. The diamonds and triangles indicate corresponding results for the reference ITDs of 200, 400, and $600 \mu s$, respectively. As a general trend, an increase of the reference ITD leads to an increase of the ITD discrimination thresholds, which is in line with results from previous studies (e.g., [Hafta et al., 1975](#); [Domnitz and Colburn, 1977](#)). The increase occurs at all center frequencies, but is more prominent at high center frequencies. For example, at 231 Hz, the threshold obtained for a reference ITD of $600 \mu s$ is 1.6 times higher than the one obtained for a reference ITD of $0 \mu s$ whereas the corresponding ratio for the center frequency of 1254 Hz is 4.5. Thus, the spectral range of the dominance region changes with the reference ITD: it is between about 750 and 1000 Hz for the reference ITD of $0 \mu s$ and lies between 250 and 600 Hz for the reference ITD of $600 \mu s$.

The results of the third experiment are shown in the right panel of Fig. 2. The thresholds obtained for the reference ITD of $0 \mu s$ (upward triangles) were replotted from Fig. 1 ($IC = 0.92$). The different symbols indicate corresponding results for the reference ITD of 200, 400, and $600 \mu s$, respectively. The effect of an increase of the reference ITD on the ITD discrimination threshold for the partially correlated signals is consistent with the results obtained with fully correlated signals (left panel), i.e., thresholds increase with increasing reference ITD whereby the increase is larger at higher frequencies. However, thresholds are generally higher for $IC = 0.92$ than for the fully correlated signals ($IC = 1$).

In summary, the obtained ITD threshold data as a function of IC and the reference ITD complement results from previous studies. ITD sensitivity was found to decrease with decreasing IC as well as with increasing reference ITD, i.e., for sound sources away from the median plane. Furthermore, the data showed a rather complex frequency dependency whereby the dominance region (i.e., the most sensitive frequency region) strongly depends both on the IC and the

reference ITD. This three-dimensional pattern of the ITD thresholds (with the dimensions center frequency, reference ITD, and IC) was described well by the proposed analytical functions, which accounted for 94% of the variance of the data. However, further investigation may improve the fit of the function to the data to better reflect the rapid roll-off of the auditory sensitivity to ITDs above about 1000 Hz ([Brughera et al., 2013](#)).

III. ITD-BASED LOCALIZATION

A. Rationale

To examine ITD-based localization performance in realistic conditions, a series of localization experiments was conducted. Broadband signals were considered in (simulated) reverberant conditions and placed at different azimuth angles. The experimental data were compared with predictions using a functional localization model which, similar to [Stern et al. \(1988\)](#), assumed an optimal integration of weighted ITD information across frequency bands. The weights of the ITD information were assumed to depend on frequency, IC , and ITD, and were derived from the ITD discrimination data presented above (Sec. II).

B. Method

1. Listeners and apparatus

Five young listeners participated in the series of four localization experiments, from which only one (i.e., subject S1) also participated in the ITD discrimination experiments described in Sec. II. The same apparatus was used in the localization and ITD discrimination experiments. The listeners responded using a computer program with a graphical interface running in MATLAB. For all statistical testing, a repeated measure analysis of variance (ANOVA) was applied using MATLAB.

2. Procedure and stimuli

The task of the listeners was to “align” the perceived lateralization of a pointer signal to that of a target signal by adjusting the ITD carried by the pointer signal. The listeners could play the target or the pointer signals at their convenience. A measurement ended when the listener decided that the lateralization of the target and pointer signals matched each other. Twelve repetitions of each condition were carried out for each listener.

Pointer and target signals consisted of nine 1-ERB-wide bands centered at 148, 231, 330, 451, 598, 776, 992, 1254, and 1572 Hz. The signals were presented at 70 dB SPL, were 300-ms long and had 5-ms long onset and offset ramps. The 2-ERB separation between two consecutive bands allowed an independent adjustment of the ITD and IC in each frequency channel. The pointer signals were fully correlated and carried a single ITD that was adjusted by the listeners with one of the three step sizes: 150, 50, or $20 \mu s$. The initial position of the pointer signal was randomly chosen between -700 and $700 \mu s$.

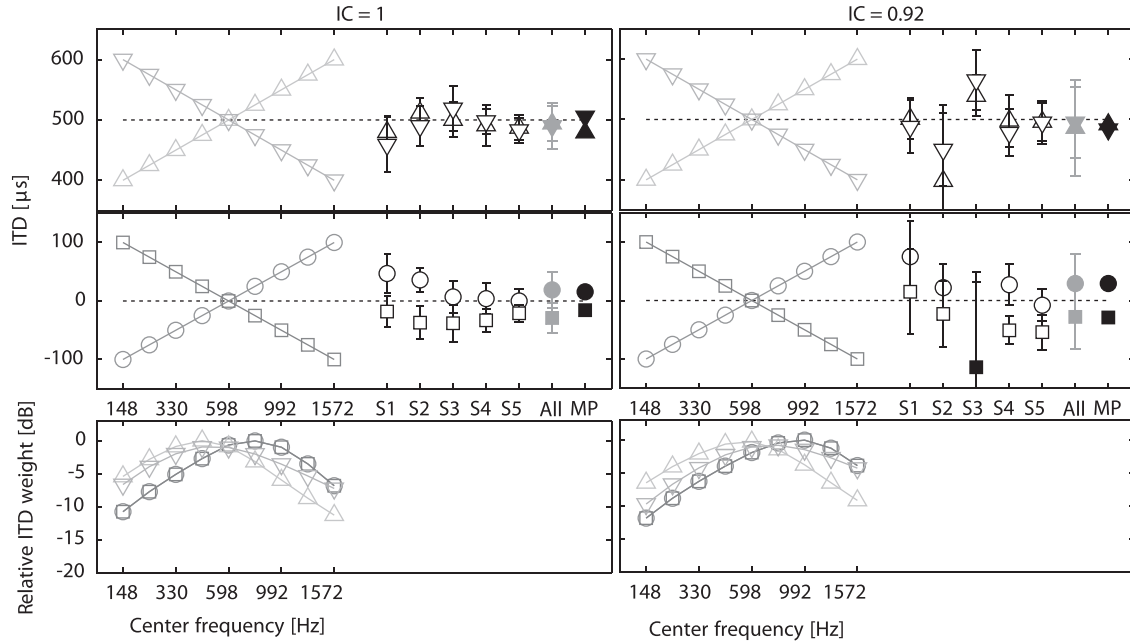


FIG. 3. Localization data obtained with broadband noise stimuli with IC = 1 (left panels) and IC = 0.92 (right panels). Top and middle panels: The light gray symbols indicate the increasing and decreasing ITD distributions as a function of center frequency. The experimental data are indicated by the black open symbols with error bars. The individual localization data for five listeners are shown by open symbols at the positions “S1–S5.” The across-listener average data are shown as gray filled symbols at the position “All.” Error bars represent the standard deviation. Model predictions are shown by the black filled symbols at the position “MP.” The bottom panels represent the normalized relative ITD weights as a function of center frequency. Weights are indicated by the same symbols as the ones used for the corresponding ITD distributions shown in the top and middle panels.

The target signals carried a different ITD in each frequency band. The ITDs were either distributed between -100 and $100 \mu\text{s}$ (frontal condition) or between 400 and $600 \mu\text{s}$ (lateral condition). The ITDs in the different frequency bands were linearly spaced within the ITD-range of the frontal or lateral condition and either increased or decreased with the increasing center frequency. The resulting four different configurations of the ITDs are indicated by the connected open gray symbols in the top and middle panels of Fig. 3. In the top panels, the increasing and decreasing ITD distributions are shown for the lateral condition, left for IC = 1 and right for IC = 0.92. The middle panels show the corresponding ITD distributions for the frontal condition. Four experiments were carried out. First, the IC was kept constant at the value of one in all frequency channels. Second, the same was done with IC = 0.92. Third, different ICs were applied in the different frequency bands whereby the IC values were linearly spaced between 0.85 and 1 (increasing IC) from the low to the high center frequency. Finally, in the fourth experiment, a linear spacing between 1 and 0.85 (i.e., a decreasing IC) was applied.

It should be noted here that the reduction of the IC of the different noise bands of the described stimuli, as well as the variation in the applied ITDs across frequency, resulted both in a widening of the perceived image of the stimuli, i.e., in an increase of the apparent source width. This was not the case for the pointer signal, which always provided a focused image due to its frequency-independent ITD as well as an IC of 1.

3. Model of spectral integration of ITDs

A functional localization model, inspired by the framework provided by Stern *et al.* (1988), was considered to

describe the data obtained in the localization of the noise signals obtained in this experiment. Similar to Stern *et al.* (1988), it was assumed that the localization of a signal can be calculated via (i) estimating the ITDs in the individual auditory frequency channels (e.g., by applying a short-term cross correlation analysis) and (ii) calculating the weighted sum over all ITDs,

$$\hat{\text{ITD}} = \sum_{i=1}^N a_i \cdot \tau_i, \quad (3)$$

where N represents the number of considered frequency channels, τ_i is the estimated ITD in frequency channel i , and a_i represents the weight of the ITD in frequency channel i . The weights are determined by the variance σ_i^2 of the internal noise in the corresponding frequency channel, normalized by the total variance averaged across the N frequency channels covered by the signal

$$a_i = \frac{\frac{1}{\sigma_i^2}}{\sum_{i=1}^N \frac{1}{\sigma_i^2}}. \quad (4)$$

This spectral weighting provides an optimal integration when the internal noise (with variance σ_i^2) is assumed to be Gaussian distributed. It was assumed here that the internal noise limiting the (ITD-based) auditory localization performance corresponds to the internal noise estimated on the basis of the ITD discrimination experiments described above (Sec. II; see also discussion in Sec. IV).

C. Results

1. Equal IC across frequency

Figure 3 shows the localization data, represented by the symbols including error bars, obtained for target signals with an IC of either 1 (left panels) and 0.92 (right panels) in all frequency bands. The top panels show the localization data for target signals in the lateral condition and the middle panels in the frontal condition. The bottom panels show the ITD weights (α_i) in the individual frequency bands of the target signals derived via Eq. (4).

Regarding the data obtained in the lateral condition with IC = 1 (top left), the average pointer ITDs, as indicated by the filled gray symbols, are essentially the same for the two ITD distributions, with values of 490 and 494 μ s, respectively. A repeated measure ANOVA did not reveal any significant effect of the ITD distribution [$F(1,4) = 0.28$, $p = 0.6251$]. This behavior is also reflected in the individual data (open symbols), even though the ITD values varied across listeners. The results obtained in the frontal condition (middle left panel) show that the average pointer ITDs obtained with the two distributions differ from each other, with values at 19 and -29μ s, respectively. This difference was also represented in the individual data and was significant [$F(1,4) = 26.21$, $p = 0.0069$]. The corresponding model predictions are indicated by the filled black symbols with the label “MP” and are consistent with their experimental data. For an IC of 1 the predicted values were 479 and 506 μ s for the lateral condition and 16 and -16μ s for the frontal condition.

For target signals with an IC of 0.92 (right panel), the listeners generally reported that the task was more difficult than with an IC of 1, which is reflected by the markedly larger error bars. Nevertheless, as in the case of IC = 1, for the lateral condition (top right panel), the average pointer ITDs were very similar for the two ITD distributions, with values of 486 and 495 μ s, despite substantially varying values across the listeners. The localization was not significantly affected by the type of ITD distribution [$F(1,4) = 0.46$, $p = 0.5361$]. For the frontal condition (right middle panel), the average pointer ITDs obtained with the two distributions were equal to 29 and -28μ s and significantly different from each other [$F(1,3) = 40.93$, $p = 0.0077$]. The listener S3 showed inconsistent results with a high variability across trials and pointer ITDs outside the range of ITDs carried by the target signal. This subject was therefore not included in the statistical analysis. The model predictions were 483 and 492 μ s for the lateral condition and 29 and -29μ s for the frontal condition. These values are consistent with their respective experimental data.

The bottom panels of Fig. 3 show the calculated (normalized) relative weights [see Eq. (4)] of the ITD information in each frequency channel for the four target signal configurations. The weights derived for the target signal configurations with ITDs in the lateral conditions (upwards and downwards pointing triangles) show a dominance of the information in the frequency channels centered at 451 (IC = 1) and 598 Hz (IC = 0.92), which carry an ITD equal to or close to the average ITD carried by the target signals. Consequently, the corresponding predictions are close to the average ITD carried by the target signal, i.e., around 500 μ s.

The model behavior is different for the frontal conditions. Due to the symmetry around 0 μ s of the ITD values, their calculated weights are equal, i.e., the squares and circles are on top of each other. Moreover, these weights show a dominance of the ITD carried at 776 Hz for IC = 1 (left) and at 992 Hz for IC = 0.92 (right). Since at these frequencies, the stimulus ITD is different from the average ITD of 0 μ s and also different between the ITD distributions (i.e., circles versus squares), the predicted ITDs are also different from the average ITD as well as between the different distributions. Consistently, these differences are largest for an IC of 0.92, which, at the frequency of maximal weight, also shows the largest differences in ITDs between ITD distributions.

2. Different ICs across frequency

The localization data obtained for target signals where the IC increased linearly between 0.85 and 1 are shown in Fig. 4 (left panels). The corresponding results for the IC decreasing between 1 and 0.85 are shown in the right panels. Regarding the conditions with increasing IC with frequency (left), it can be seen that the frequency weights are very large for noise bands at and above 992 Hz. As a consequence, the predictions for target signals carrying increasing and decreasing ITDs are far apart from one another for both the frontal and lateral conditions. For the frontal condition (middle panel), the predictions are 42 and -42μ s with a difference of 84 μ s. These predictions are very well in line with the average experimentally obtained ITDs, with values at 28 and -52μ s, i.e., a difference of 80 μ s. The localization was significantly affected by the type of the ITD distribution [$F(1,4) = 40.93$, $p = 0.01$]. For the target signals in the lateral condition (top-left panel), the predictions also show a relatively large difference between the predictions for the two ITD distributions, 466 and 512 μ s. The average pointer ITDs, although further apart than in the experiments with fixed ICs (Fig. 3), are less different from one another, with values of 478 and 508 μ s. The localization was just not affected by the type of ITD distribution [$F(1,4) = 6.46$, $p = 0.0639$] due to the rather large variability in the individual data. For listeners S1 and S2, a clear difference can be observed between the pointer ITDs for the two ITD distributions with 62 and 53 μ s, which is well in line with the difference in the model predictions of 46 μ s. In contrast, for listeners S3–S5, the difference was virtually zero.

In the case of the IC decreasing from 1 to 0.85 between low and high frequencies (left panels), the estimated ITD weights are more homogeneous across frequency than in the other experiments, and show even a slight low-frequency dominance for the lateral conditions. As a consequence, the predicted localization obtained with the two different ITD distributions for the frontal condition (middle right panel) was close to the average ITD of 0 μ s, with values of -8 and 8 μ s. These predictions are very similar to the average pointer ITDs of -21 and 13 μ s, which showed a small but significant effect of ITD distribution [$F(1,4) = 46.16$, $p = 0.0025$].

For the target signals in the lateral condition (top-right panel in Fig. 4), despite the large variability across listeners,

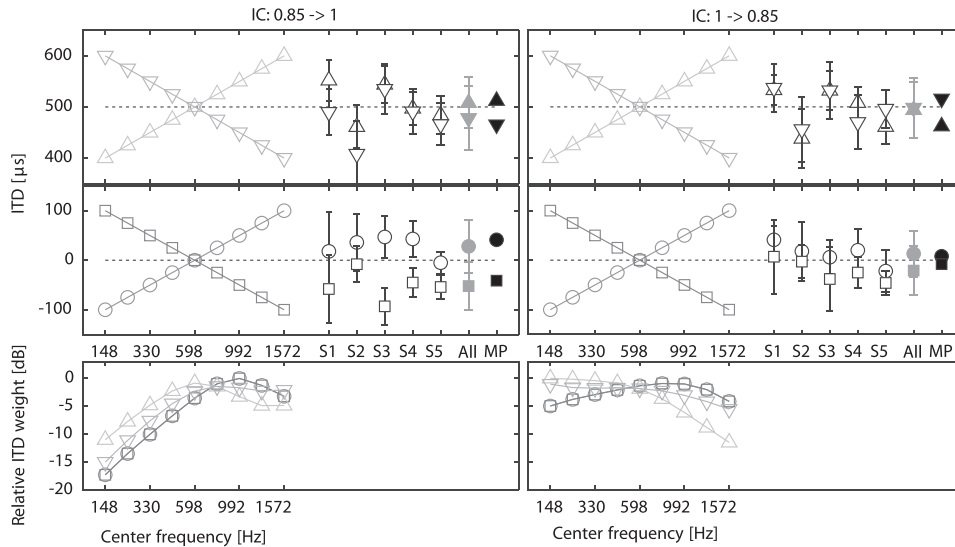


FIG. 4. Similar as in Fig. 3 but with IC values that were linearly spaced from 0.85 to 1 (left panels) or from 1 to 0.85 (right panels) from low to high center frequencies of the noise bands.

the average pointer ITDs for the two distributions are very close to one another, 494 and 498 μs , and not significantly different [$F(1,4) = 0.1$, $p = 0.7695$]. These average data are closer to each other than suggested by the model predictions, which are 462 and 516 μs .

D. Discussion

Overall, the experimental data could be reasonably well accounted for by the functional localization model. In a number of stimulus conditions, it seems that the localization could be the result of a simple average of the ITDs carried by the target signals. The calculated ITD weights suggest, however, that this is not generally the case. For example, in the lateral conditions with constant IC (Fig. 3, top panels), the average pointer ITDs of 490 and 494 μs ($\text{IC} = 1$) and 486 and 495 μs ($\text{IC} = 0.92$), were close to the average ITD of 500 μs as a result of the strong dominance of the ITD carried in a rather narrow frequency channel around 500 Hz, in which the target signal had an ITD that was coincidentally close to 500 μs . In contrast, the pointer ITDs in the frontal condition with decreasing ICs of -21 and $13 \mu\text{s}$ (Fig. 4, middle-right panel) were close to the average ITD of $0 \mu\text{s}$ because of the rather homogeneous ITD weighting across frequency, which basically realized an averaging operation.

The good agreement between the measured and predicted localization data suggests that the auditory system integrates ITDs “optimally” across frequencies, as described by Eq. (4). However, one may consider an alternative hypothesis, in which no spectral integration was assumed and only the frequency channel in which the ITD is the most salient would be considered. This alternative hypothesis was also tested in the framework of the model. In a “single-channel” version of the model, the weight of the most salient channel was set to 1 and all other channel weights were set to 0. The predictions of the single-channel model were in good agreement with the localization data for three of the experimental conditions (conditions with constant IC and for ICs increasing with increasing frequency), although the overall error was larger than for the “multi-channel” model.

This is due to the fact that, in these conditions, the calculated ITD weights in the multi-channel model show dominance in a relatively narrow frequency range. However, in the condition with decreasing ICs with increasing frequency (Fig. 4, right panels), where the predicted ITD weights are distributed more homogeneously across frequencies, the single-channel model provided results that differed more strongly from the measured data. For the frontal condition, for instance, the average pointer ITD for the two ITD distributions was -21 and $13 \mu\text{s}$, whereby the corresponding predictions were -11 and $11 \mu\text{s}$ for the multi-channel model and -50 and $50 \mu\text{s}$ for the single-channel model. Likewise, for the lateral condition, the average pointer ITDs were 494 and 498 μs , and the corresponding predictions were 477 and 504 μs for the multi-channel model and 500 and 550 μs for the single-channel model. Thus, although the single-channel model can successfully describe a large part of the measured ITD localization data, the multi-channel model can additionally account for the conditions where the single-channel model predictions deviate significantly from the average data.

IV. OVERALL DISCUSSION

A. Localization weights

The results of this study confirm the general conclusions of previous studies that, when lateralizing broadband stimuli, the auditory system applies an optimally weighted integration of ITD information across frequency channels. Whereas in previous studies mainly the effect of stimulus frequency and target ITD were considered within the applied weights (Stern *et al.*, 1988; Shackleton *et al.*, 1992) the present study additionally included the effect of a decrease in IC (as introduced by room reverberation). Moreover, the weights were derived directly from an extensive set of measured ITD thresholds (Sec. II), which is conceptually similar to the approaches described by Domnitz and Colburn (1977) and Stecker and Bibee (2014), but differs from the above studies where weighting functions were derived rather heuristically.

The weights derived in this study [Eq. (4)] are shown in the left three panels of Fig. 5, expressed in dB, for ICs of 1

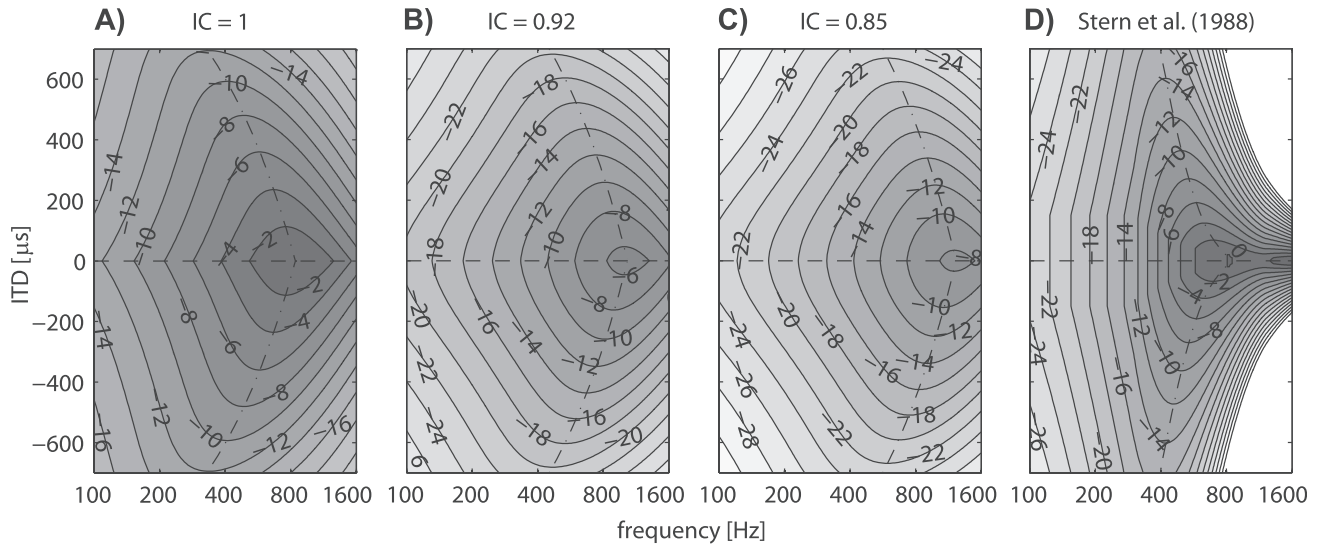


FIG. 5. Contour lines of the normalized absolute ITD weights expressed in dB as a function of frequency and ITD. (A), (B), and (C) show the weights described by Eq. (4) for the different IC values displayed above each panel. For comparison purposes, (D) shows the weights proposed by Stern *et al.* (1988), as further described in the text. Dashed-dotted lines indicate the maximum of the weights as a function of ITD. The weights in (D) were truncated below -32 dB, the minimum weight observed in (A)–(C).

[Fig. 5(A)], 0.92 [Fig. 5(B)], and 0.85 [Fig. 5(C)], for frequencies between 100 and 1600 Hz and for ITD values between -700 and 700 μ s. The weights represented in the three panels were normalized here to the largest weight which was found for a frequency of 845 Hz, an ITD of 0 μ s, and an IC of 1. For IC = 0.92, the largest weight was -5.4 dB and for IC = 0.85, the largest weight was -7.6 dB. The weights reflect a nonlinear dependency on frequency and ITD. The patterns are, however, fairly similar across IC values, with a shift of the overall pattern toward higher frequencies for smaller IC values. For each IC value, the largest weight was found for an ITD of 0 μ s and for frequencies of 850 Hz (for IC = 1), 1051 Hz (for IC = 0.92), and 1252 Hz (for IC = 0.85). The weights decrease strongly for frequencies and ITD values away from the point of maximal weight, which is particularly pronounced toward higher frequencies. As a result of this behavior, the weights exhibit a dominance region (i.e., a frequency region of maximal ITD sensitivity) that depends on both IC and ITD, as indicated in the figure by the dashed-dotted lines. This is different from Raatgever (1980), who reported an emphasis of the ITD information at around 540 Hz. In the present study, the frequency of the maximum shifts downwards with increasing ITD and upwards with decreasing IC. For example, at an ITD of 0 μ s and an IC of 1, the maximum (normalized) weight (0 dB) is at a frequency of about 850 Hz. Changing the ITD to 600 μ s results in a reduced maximum weight by -10 dB and a shift to 400 Hz. Similarly, changing the IC to 0.85 (and keeping the ITD at 0 μ s) results in a reduced weight of the maximum weight by -7.6 dB and a shift to 1252 Hz.

To compare the weights derived in this study with the ones described by Stern *et al.* (1988), their results are shown in Fig. 5(D). These weights were calculated by setting the straightness parameter σ_i^2 to zero (i.e., assuming tonal stimuli), which resulted in a weighting function (in dB) of $10 \cdot \log_{10}[p(\tau f_0) \cdot q(f_0)]$, with the functions $p(\tau f_0)$ and $q(f_0)$ provided in Stern *et al.* (1988, p. 160). The weights were

normalized to their maximum value, which occurred at an ITD of 0 μ s and a frequency of 827 Hz. The weights described by Stern *et al.* (1988) exhibit a similar qualitative behavior as the weights derived in the present study but reflect an increased dynamic range: The dependency on frequency is more pronounced and the decay with increasing ITD is substantially steeper, particularly at high frequencies. Whereas the dynamic range shown in any of the panels (A)–(C) in Fig. 5 is about 20 dB, and about 32 dB across the three panels, the dynamic range of the weights of Stern *et al.* (1988) is far larger, and was therefore truncated in panel (D) below a weight of -32 dB. This increased dynamic range in Stern *et al.* (1988) also resulted in an ITD-dependent dominance region that is much narrower than shown in panels (A)–(C), in particular at high frequencies. Since Stern *et al.* (1988) did not consider the impact of IC on localization, it is not considered in their weights.

The increased dynamic range found for the ITD weights of Stern *et al.* (1988), and in particular the faster decay with increasing frequency, may partly be explained by the difference in bandwidth of the applied stimuli. Whereas in Sec. II the weights were measured using narrowband (1-ERB wide) noise, in Stern *et al.* (1988) (for $\sigma_i^2 = 0$) tonal signals were assumed. The increased stimulus bandwidth may have introduced a spectral smoothing to the ITD weighting-functions and thereby reduced the spectral variations, including the frequency roll-off. The difference may be slightly reduced by increasing the straightness parameter σ_i^2 when calculating the weights from Stern *et al.* (1988).

To the best knowledge of the authors, no data exist in the literature that can be directly compared to the IC dependency of the derived localization weights. Figure 5 (panels A to C) illustrates that with decreasing coherence the overall weights decrease and the effect of frequency as well as reference ITD on the weights is less pronounced (i.e., the weighting functions become more compressed). Moreover, the maximum of the ITD-dependent dominance region shifts

toward higher frequencies for all considered ITDs. The general reduction of the localization weights with decreasing IC is in qualitative agreement with [Faller and Merimaa \(2004\)](#), who argued that only ITDs with an IC above a certain threshold contribute to localization.

B. Physiological considerations

The relative weighting of ITDs and the influence of room reverberation on auditory localization has been investigated in a few physiological studies. The first attempt to formalize ITD processing was the conceptual “coincidence detectors” proposed in [Jeffress \(1948\)](#). Jeffress assumed that ITDs were internally coded by detectors sensitive to a specific ITD as well as frequency. Considering physiological knowledge of his time, Jeffress suggested that detectors tuned to large ITDs would require a longer path between the ears and would therefore be less numerous, suggesting that the sensitivity to ITD-changes decreases with increasing ITD. The concept of an azimuthal space (and frequency) map is in line with the weights derived in this study. As shown in [Fig. 5](#), the relative weights not only depend on ITD, but also on frequency, and the dependence on the ITD varies greatly with frequency. However, an increasing body of more recent research argues against the existence of an azimuthal space map within the human auditory system as inferred by the Jeffress model (e.g., see the review by [Grothe et al., 2010](#)). In this regard, it should be emphasized that even though the ITD-weighting function inherent in the present localization model applies an azimuthal space-frequency map, the model does not rely on the existence of such map within the auditory system. The map mainly reflects the stimulus manipulations applied in the above experiments and simplifies the mathematical framework of the localization model, but does not inform about the underlying auditory processes involved in ITD coding. The effect of room reverberation on the neural coding of low-frequency ITDs has been measured in the midbrain of anesthetized cats ([Devore et al., 2009](#)). Among other aspects, it was observed that room reverberation degrades the directional sensitivity of single neurons, in particular, in the later or steady-state portion of the signal. This is in general agreement with the present finding (see [Fig. 5](#)) that a decrease in IC (due to reverberation) results in a reduction of the ITD weighting.

C. Implications for existing binaural models

It should be noted that the weighted-image model solely considers the location of the maxima of the (long-term) cross-correlation function and thus, the spectral weighting and integration of the ITD information is completely decoupled from the actual realization of the cross-correlation function. Hence, such conceptual approach does not take the height of the cross-correlation function into account, which provides a direct measure of the IC, nor does it describe other (often non-linear) mechanisms that are inherent in most existing binaural models (e.g., [Colburn, 1977](#); [Cai et al., 1998](#); [Lindemann, 1986](#); [Gaik, 1993](#); [Breebaart et al., 2001](#); [Dietz et al., 2011](#); [Kayser et al., 2015](#)) and may affect

the weighting of ITD information within a subsequent spectral integration process. The contra-lateral inhibition mechanism proposed by [Lindemann \(1986\)](#) and extended by [Gaik \(1993\)](#), for example, has a strong non-linear effect on the amplitude of the estimated binaural (cross-correlation) function that is highly dependent on the reference-ITD, ILD, and IC as well as the history of the signals at the two ears. However, due to the complicated and non-linear behavior of these binaural models, the realized ITD-weighting that is relevant to a subsequent spectral integration mechanism is not known. Moreover, no conclusive psychoacoustical data set is available that allows the verification (and optimization) of the ITD-weighting (or sensitivity) inherent in these binaural models, in particular with respect to changes in IC. In this regard, the derivation of the extensive data set of ITD thresholds ([Sec. II](#)) as well as the corresponding ITD weights for auditory localization of broadband signals ([Sec. III](#)) may be valuable for the development and evaluation of signal-driven auditory localization models.

D. Limitations and perspectives

The current study solely considered ITDs carried by the temporal fine-structure of the steady-state portion of band-limited noise with varying IC at low frequencies (i.e., at frequencies below 1.5 kHz). However, the auditory system also utilizes ILDs to localize sounds as well as ITDs carried by the signal’s envelope. Moreover, when signals are presented in rooms, the IC varies over the time course of the signal, typically providing high IC values at (echo-free) signal onsets and reduced values in later, steady-state portions of the signal. In such case, auditory localization is typically most sensitive to the early portion of the signal and puts less weight on the later portion of the signal (e.g., [Devore et al., 2009](#); [Devore and Delgutte, 2010](#); [Stecker and Moore, 2018](#)). Even though these aspects were not considered here, the presented methods may be extended to measure the (relative) weights across the different cues as well as over the time course of a reverberant signal with time- (and frequency-) varying IC. Such research would complement other relevant studies that applied amplitude modulated stimuli to determine the temporal weighting of interaural cues (e.g., [Dietz et al., 2013](#); [Stecker and Bibee, 2014](#); [Hu et al., 2017](#)).

Moreover, rather artificial stimuli were applied here to systematically study the effect of specific signal parameters (i.e., IC, frequency, and reference ITD) on auditory localization, but such stimuli are rarely encountered in the real world. Hence, important signal properties as well as auditory phenomena that can be observed in the real world were not considered. Realistic stimuli, such as speech, contain temporal modulations as well as distinct spectral features that change over time. Such stimuli typically exhibit onsets and modulations that are correlated across frequency and provide pitch information. The auditory processes that utilize these acoustic features were not considered here but may be considered in future studies.

Finally, the observed ITD sensitivity showed a substantial variation across listeners, but only average data were further evaluated. With respect to the localization data, some listeners showed a bias across all experimental conditions, with listener S2, for instance, consistently underestimating the laterality of the stimuli and listener 3 overestimating it. Across-listener differences may shed light on the different spatial cue weighting as well as localization strategies that are applied by the different listeners, which may become even more apparent when additional localization cues are taken into account (i.e., envelope ITDs and ILDs) or more realistic stimuli are considered. Future studies may examine these individual differences across listeners in ITD sensitivity (or cue weighting) as well as localization, and the effects of hearing loss.

V. SUMMARY AND CONCLUSION

Auditory sensitivity of the ITD carried by the signal's temporal fine-structure was measured in four normal-hearing listeners as a function of frequency, reference ITD, and IC using critical-band wide noise. The resulting average ITD thresholds were approximated by a set of analytical functions and localization weights were derived using concepts from signal detection theory. The weights were then applied in a simple localization model that was proposed to describe the weighted integration of ITDs across the frequency. To verify this model, experiments were conducted that assessed the perceived lateralization of low-frequency noise that consisted of nine critical-band wide noise bands, which were separated in frequency to minimize spectral overlap and differed by their ITD as well as IC. The resulting data were compared to predictions obtained with the proposed localization model. The good agreement observed between the localization data and the model predictions supports the hypothesis that the auditory system performs a weighted integration of ITDs across frequency to localize a broadband sound source. The applied experimental methods and modeling concepts may help design future psychoacoustical experiments that evaluate the impact of additional signal features on localization, including the temporal behavior of the IC in rooms, ILDs, and signal envelope-based ITDs. The derivation of the extensive ITD threshold data as well as the corresponding localization weights could be useful for the development and evaluation of signal-driven auditory models to predict auditory localization of complex stimuli in reverberant environments.

ACKNOWLEDGMENTS

Part of the data as well as the first version of the model presented here was presented at the ICA/ASA Meeting on Acoustics, Montréal 2013 (Le Goff *et al.*, 2013b). This study was supported by the Danish Research Foundation (DK) as well as the HEARing CRC, established and supported under the Cooperative Research Centres Program—an initiative of the Australian Government.

APPENDIX A

To provide an analytical approximation for the ITD thresholds measured in Sec. II, which is required for the localization model described in Sec. III, Eq. (2) needs to be applied to Eq. (1) and transformed into

$$\Delta_\tau(f_0, \tau, \rho) = \sqrt{\sigma_{\text{int}}^2(f_0, \tau) + \sigma_{\text{ext}}^2(f_0, \tau, \rho)} \cdot d', \quad (\text{A1})$$

where Δ_τ is the fitted continuous function of the ITD thresholds as a function of frequency f_0 , ρ is the interaural correlation, and τ is the relevant ITD given by

$$\tau = \tau_{\text{ref}} - \frac{\Delta_\tau}{2}. \quad (\text{A2})$$

Within Eq. (A2), it is assumed that in an ITD discrimination experiment, the variances of the noise limiting the representation of the ITD is estimated half-way between the reference ITD, τ_{ref} , and the target ITD defined as $\tau_{\text{target}} = \tau_{\text{ref}} - \Delta_\tau$ (see Sec. II B for details). The internal and external variances σ_{int}^2 and σ_{ext}^2 in Eq. (A1) are arbitrarily defined as

$$\sigma_{\text{int}}^2(f_0, \tau) = g_{\text{int}}(\tau) \cdot \left(\frac{1}{B_{\text{ERB}}(f_0) \cdot f_0} \right) \cdot \left(1 + \left(\frac{f_0}{f_{\text{int}}(\tau)} \right)^{N(\tau)} \right), \quad (\text{A3})$$

$$\sigma_{\text{ext}}^2(f_0, \tau, \rho) = g_{\text{ext}}(\tau, \rho) \cdot \left(\frac{1}{B_{\text{ERB}}(f_0) \cdot f_0} \right) \cdot \left(1 + \left(\frac{f_0}{f_{\text{ext}}(\tau, \rho)} \right)^{N(\tau)} \right), \quad (\text{A4})$$

where $B_{\text{ERB}}(f_0)$ is the ERB of an auditory filter at center frequency f_0 , which according to Patterson *et al.* (1988) is given by

$$B_{\text{ERB}}(f_0) = 24.7 \cdot (0.00437 \cdot f_0) + 1. \quad (\text{A5})$$

The frequency f_0 and bandwidth B_{ERB} are given in Hertz and the interaural delay τ is given in microseconds. Equations (A3) and (A4) describe band-stop shaped variances with an overall sensitivity defined by g_{int} and g_{ext} as given in Eqs. (A6) and (A7). The second term in Eqs. (A3) and (A4) describes a low-frequency roll-off that mainly reflects an observer with constant-phase sensitivity, as described by the term $1/f_0$. The term $1/B_{\text{ERB}}(f_0)$ refers to the observation that the variance of a (auditory) bandpass filtered noise is inversely proportional to its bandwidth. The last term in Eqs. (A3) and (A4) introduces a high-pass characteristic with cut-off frequencies f_{int} and f_{ext} as well as order N , which are all dependent on the interaural delay τ and for f_{ext} also on the IC ρ . This high-pass characteristic mainly refers to the limitations of the high-frequency coding observed in the inner hair-cells as well as the subsequent neural pathways of the binaural auditory system. These functions are given by

$$g_{\text{int}}(\tau) = 0.06 \cdot \tau + 22 \times 10^{-6}, \quad (\text{A6})$$

$$g_{\text{ext}}(\tau, \rho) = \frac{(3500 \cdot \tau)^2 + 1}{500} \cdot \left[(\rho^2 - 1.07)^2 - 4.9 \times 10^{-3} \right], \quad (\text{A7})$$

$$f_{\text{int}}(\tau) = -0.85 \times 10^6 \cdot \tau + 950, \quad (\text{A8})$$

$$f_{\text{ext}}(\tau, \rho) = f_{\text{int}}(\tau) \cdot (-2.6 \cdot \rho^2 + 3.6), \quad (\text{A9})$$

$$N(\tau) = -(1700 \cdot \tau)^2 + 5, \quad (\text{A10})$$

whereas Eqs. (A1)–(A5) are inspired by either auditory or signal processing concepts, Eqs. (A6)–(A10) have no direct physical or auditory relevance, except for the dependency of the IC ρ within Eq. (A7). The coefficients in Eqs. (A6)–(A10) were numerically fitted in MATLAB by minimizing the mean squared error between the experimental data measured in Sec. II and the corresponding analytical approximations given in Eqs. (A1)–(A4).

APPENDIX B

The individual and mean ITD thresholds measured in experiments 1–3 (Sec. II) are summarized in Tables I and II. The mean values were derived from the individual ITD thresholds applying a logarithmic transformation. Note that the four subjects that participated in experiment 1 (Table I) are different from the four subjects that participated in experiments 2 and 3 (Table II), except for subject S1 who participated in all three experiments.

TABLE I. Individual and mean ITD thresholds in microseconds for experiment 1.

Subject	IC	ITD _{ref} (μ s)	Center frequency of narrowband noise (Hz)								
			148	231	330	451	598	776	992	1254	1572
S1	1.00	0	40.0	44.0	30.0	23.1	23.4	18.7	18.3	36.2	44.9
	0.97	0		73.8		57.6		35.6		40.1	
	0.92	0		131.7		68.6		44.2		37.5	
	0.85	0		176.1		122.0		76.6		56.7	
S2	1.00	0	69.9	36.6	29.5	32.9	27.0	26.3	21.4	21.9	29.5
	0.97	0		101.6		52.1		23.7		24.4	
	0.92	0		222.6		58.8		35.0		38.4	
	0.85	0		276.6		118.2		67.3		37.6	
S3	1.00	0	93.0	75.3	63.3	38.1	37.0	23.4	30.6	30.6	44.1
	0.97	0		124.0		59.8		31.8		34.4	
	0.92	0		252.9		86.4		52.0		46.8	
	0.85	0		295.8		122.0		74.0		66.1	
S4	1.00	0	100.3	69.7	41.6	43.2	28.4	23.8	23.4	30.0	30.6
	0.97	0		110.4		44.1		40.0		29.5	
	0.92	0		182.9		68.6		52.5		47.7	
	0.85	0		225.1		136.9		63.5		72.7	
Mean	1.00	0	71.4	53.9	39.1	33.4	28.5	22.9		29.2	36.5
	0.97	0		100.6		53.1		32.2		31.6	
	0.92	0		191.9		69.9		45.3		42.3	
	0.85	0		238.6		124.6		70.2		56.6	

TABLE II. Individual and mean ITD thresholds in microseconds for experiment 2 and 3.

Subject	IC	ITD _{ref} (μ s)	Center frequency of narrowband noise (Hz)								
			148	231	330	451	598	776	992	1254	1572
S1	1.00	200	39.5	27.9	28.4	36.7	25.2	22.5	23.8	20.7	46.9
	1.00	400	61.7	49.6	49.6	37.9	35.7	40.0	33.2	44.8	66.1
	1.00	600	75.5	62.5	56.6	40.6	49.6	56.5	48.3	60.7	92.2
	0.92	200		209.5		77.7		50.5		33.3	
	0.92	400		169.0		99.4		70.5		57.5	
	0.92	600		169.0		130.3		64.3		102.0	
S5	1.00	200	100.7	79.8	57.7	56.5	43.2	41.6	37.0	43.2	71.0
	1.00	400	100.7	94.7	67.1	51.3	44.9	51.3	54.5	93.3	104.6
	1.00	600	117.5	84.7	71.2	73.9	64.6	64.0	133.8	108.7	156.1
	0.92	200		145.0		74.1		84.8		53.5	
	0.92	400		150.5		84.7		67.3		156.3	
	0.92	600		233.9		144.8		200.3		238.1	
S6	1.00	200	136.5	75.3	49.4	53.3	37.8	31.2	55.5	56.5	58.8
	1.00	400	139.5	102.3	79.8	51.3	55.4	71.0	100.4	129.7	171.8
	1.00	600	182.3	114.9	82.9	79.8	78.2	128.7	144.5	165.7	178.8
	0.92	200		229.3		77.0		78.4		98.8	
	0.92	400		200.7		91.6		150.6		225.3	
	0.92	600		225.2		172.1		181.1		283.0	
S7	1.00	200	168.8	68.4	61.0	55.4	49.4	41.8	51.3	79.7	110.5
	1.00	400	243.1	114.8	63.3	86.1	81.3	133.8	112.8	150.4	162.4
	1.00	600	225.2	96.9	96.9	136.3	156.0	215.8	226.0	189.6	197.0
	0.92	200		153.5		115.2		89.7		115.1	
	0.92	400		156.4		147.7		175.6		189.7	
	0.92	600		289.1		216.7		317.9		466.8	
Mean	1.00	200	97.8	58.2	47.1	49.7	37.8	33.2	39.8	44.8	68.2
	1.00	400	120.5	86.2	64.0	54.1	51.8	66.5	67.3	95.0	117.9
	1.00	600	138.1	87.6	75.4	75.6	79.1	100.1	120.5	120.0	150.1
	0.92	200		180.8		84.5		74.1		67.1	
	0.92	400		168.1		103.3		105.8		140.0	
	0.92	600		225.2		162.9		165.0		238.0	

- Bernstein, L. R., and Trahiotis, C. (2008). "Binaural signal detection, overall masking level, and masker interaural correlation: Revisiting the internal noise hypothesis," *J. Acoust. Soc. Am.* **124**, 3850–3860.
- Blauert, J. (1986). "Auditory spaciousness: Some further psychoacoustic analyses," *J. Acoust. Soc. Am.* **80**, 533–542.
- Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*, Rev. ed. (MIT Press), pp. 203–237.
- Breebaart, J., van de Par, S., and Kohlrausch, A. (2001). "Binaural processing model based on contralateral inhibition. I. Model structure," *J. Acoust. Soc. Am.* **110**, 1074.
- Brughera, A., Dunai, L., and Hartmann, W. M. (2013). "Human interaural time difference thresholds for sine tones: The high-frequency limit," *J. Acoust. Soc. Am.* **133**, 2839–2855.
- Cai, H., Carney, L. H., and Colburn, H. S. (1998). "A model for binaural response properties of inferior colliculus neurons. I. A model with interaural time difference sensitive excitatory and inhibitory inputs," *J. Acoust. Soc. Am.* **103**, 475–493.
- Colburn, H. S. (1977). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," *J. Acoust. Soc. Am.* **61**, 525–533.
- Devore, S., and Delgutte, B. (2010). "Effects of reverberation on the directional sensitivity of auditory neurons across the tonotopic axis: Influences of interaural time and level differences," *J. Neurosci.* **30**, 7826–7837.
- Devore, S., Ihlefeld, A., Hancock, K., Shinn-Cunningham, B., and Delgutte, B. (2009). "Accurate sound localization in reverberant environments is mediated by robust encoding of spatial cues in the auditory midbrain," *Neuron* **62**, 123–134.

- Dietz, M., Ewert, S. D., and Hohmann, V. (2011). "Auditory model based direction estimation of concurrent speakers from binaural signals," *Speech Commun.* **53**, 592–605.
- Dietz, M., Marquardt, T., Salminen, N. H., and McAlpine, D. (2013). "Emphasis of spatial cues in the temporal fine structure during the rising segments of amplitude-modulated sounds," *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15151–15156.
- Domnitz, R., and Colburn, H. (1977). "Lateral position and interaural discrimination," *J. Acoust. Soc. Am.* **61**, 1586–1598.
- Durlach, N. I. (1963). "Equalization and cancellation theory of binaural masking-level difference," *J. Acoust. Soc. Am.* **35**, 1206–1218.
- Faller, C., and Merimaa, J. (2004). "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," *J. Acoust. Soc. Am.* **116**, 3075–3089.
- Gaik, W. (1993). "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling," *J. Acoust. Soc. Am.* **94**, 98–110.
- Glasberg, B., and Moore, B. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.
- Green, D., and Swets, J. (1966). *Signal Detection Theory and Psychophysics* (John Wiley & Sons Ltd., New York).
- Grothe, B., Pecka, M., and McAlpine, D. (2010). "Mechanisms of sound localization in mammals," *Physiol. Rev.* **90**, 983–1012.
- Hacker, M. J., and Ratcliff, R. (1979). "A revised table of d for M-alternative forced choice," *Percept. Psychophys.* **26**, 168–170.
- Haft, E. R., De Maio, J., and Hellman, W. S. (1975). "Difference thresholds for interaural delay," *J. Acoust. Soc. Am.* **57**, 181–187.
- Hartmann, W., Rakerd, B., and Koller, A. (2005). "Binaural coherence in rooms," *Acta Acust. Acust.* **91**, 451–462.
- Hartmann, W. M., and Cho, Y. J. (2011). "Generating partially correlated noise—a comparison of methods," *J. Acoust. Soc. Am.* **130**, 292–301.
- Hu, H., Ewert, S. D., McAlpine, D., and Dietz, M. (2017). "Differences in the temporal course of interaural time difference sensitivity between acoustic and electric hearing in amplitude modulated stimuli," *J. Acoust. Soc. Am.* **141**, 1862–1873.
- ISO 3382-1 (2009). "Acoustics—Measurement of room acoustic parameters—Part1: Performance spaces." Annex B (International Organization for Standardization, Geneva, Switzerland).
- Jeffress, L., Blodgett, H., and Deatherage, B. (1962). "Effect of interaural correlation on the precision of centering a noise," *J. Acoust. Soc. Am.* **34**, 1122–1123.
- Jeffress, L. A. (1948). "A place theory of sound localization," *J. Comp. Physiol. Psychol.* **41**, 35–39.
- Kayser, H., Hohmann, V., Ewert, S. D., Kollmeier, B., and Anemüller, J. (2015). "Robust auditory localization using probabilistic inference and coherence-based weighting of interaural cues," *J. Acoust. Soc. Am.* **138**, 2635–2648.
- Klumpp, R., and Eady, H. (1956). "Some measurements of interaural time difference thresholds," *J. Acoust. Soc. Am.* **28**, 859–860.
- Kopčo, N., and Shinn-Cunningham, B. (2002). "Auditory localization in rooms: Acoustic analysis and behavior," in *the 32nd International Acoustical Conference—EAA Symposium*, Slovakia (September 10–12), pp. 109–112.
- Kuttruff, H. (2000). *Room Acoustics*, 4th ed. (Taylor & Francis, London).
- Le Goff, N., Buchholz, J. M., and Dau, T. (2013a). "Modeling horizontal localization of complex sounds in the impaired and aided impaired auditory system," in *the Technology of Binaural Listening*, edited by J. Blauert (Springer Berlin Heidelberg, Berlin, Heidelberg), pp. 121–144.
- Le Goff, N., Buchholz, J. M., and Dau, T. (2013b). "Spectral integration of interaural time differences in auditory localization," in *Proceedings of Meetings on Acoustics ICA/ASA 2013*, Vol. 19, p. 050160.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Lindemann, W. (1986). "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals," *J. Acoust. Soc. Am.* **80**, 1608–1622.
- Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (1999). "The precedence effect," *J. Acoust. Soc. Am.* **106**, 1633–1654.
- Moore, B. J. C. (2012). *An Introduction to the Psychology of Hearing*, 6th ed. (Emerald Press, United Kingdom).
- Patterson, R., Holdsworth, J., Nimmo-Smith, I., and Rice, P. (1988). "SVOS final report: The auditory filterbank," APU Report.
- Raatgever, J. (1980). "On the binaural processing of stimuli with different interaural phase relations," Ph.D. thesis, Delft Technical University.
- Rakerd, B., and Hartmann, W. (2005). "Localization of noise in a reverberant environment," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigne, S. McAdams, and L. Collet (Springer-Verlag, New York), pp. 348–354.
- Rakerd, B., and Hartmann, W. M. (2010). "Localization of sound in rooms. V. Binaural coherence and human sensitivity to interaural time differences in noise," *J. Acoust. Soc. Am.* **128**, 3052–3063.
- Shackleton, T., Meddis, R., and Hewitt, M. (1992). "Across frequency integration in a model of lateralization," *J. Acoust. Soc. Am.* **91**, 2276–2279.
- Stecker, G. C., and Bibee, J. M. (2014). "Nonuniform temporal weighting of interaural time differences in 500 Hz tones," *J. Acoust. Soc. Am.* **135**, 3541–3547.
- Stecker, G. C., and Moore, T. M. (20138). "Reverberation enhances onset dominance in sound localization," *J. Acoust. Soc. Am.* **134**, 786–793.
- Stern, R. M., and Shear, C. D. (1998). "Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay," *J. Acoust. Soc. Am.* **100**, 2278–2288.
- Stern, R. M., Zeiberg, A., and Trahiotis, C. (1988). "Lateralization of complex binaural stimuli: A weighted-image model," *J. Acoust. Soc. Am.* **84**, 156–165.
- Trahiotis, C., and Stern, R. M. (1989). "Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase," *J. Acoust. Soc. Am.* **86**, 1285–1293.
- Wallach, H., Newman, E., and Rosenzweig, M. (1949). "The precedence effect in sound localization," *Am. J. Psychol.* **62**, 315–336.
- Westermann, A., Buchholz, J. M., and Dau, T. (2013). "Binaural dereverberation based on interaural coherence histograms," *J. Acoust. Soc. Am.* **133**, 2767–2777.
- Zwislocki, J., and Feldman, R. (1956). "Just noticeable differences in dichotic phase," *J. Acoust. Soc. Am.* **28**, 860–864.